CSAE Working Paper WPS/2022-01

When are large female-led firms more resilient against shocks?

Learnings from Indian enterprises during Covid-19 with Diff-in-Diff and Causal Forests

Merlin Stein[‡]

November 2021

Abstract

In which kind of companies did the prevalence of women on corporate boards matter during the first wave of Covid-19? 2500 large Indian firms, of which only 78% initially complied with an exogenous gender quota enable a quantitative evaluation. By comparing their quarterly revenues with a Difference-in-Differences analysis, this research initially finds a significantly positive relationship between the compliance with the one-female-board-member policy and the change in revenues during the first economic shock of the Covid-19 crisis in 2020. A Triple-Difference and Causal Forest analysis indicates that this is likely endogenously driven by (self-)selection based on sectors, capital dynamics, size, independence of directors and further firm characteristics. There is no simple association of female directors and crisis revenues: The spectrum encompasses a mix of firms with a positive, a neutral and a negative association. With rare causal context for evaluating board gender diversity or other corporate governance and ESG dynamics, this piece of research illustrates the value and limitations of applying adjusted Random Forests to overcome linearity and dimensionality limitations for deeply understanding heterogeneities-within-heterogeneities as indicators of relevant distinctions.

[‡]Independent Researcher, Formerly University of Oxford and University of Tuebingen. stein.merlin@gmail.com.
I am grateful for many thought-restructuring discussions with Christopher Woodruff, and valuable feedback from Muhammad Meki, Christian Meyer and Ana Vaz. A big thank you for helpful comments, Stefanie Almeida, Razan Amine, Tom Carlowitz, Tim Deisemann, Stefan Glaisner, Louis-Maël Jean, Julia Lawson-Johns and Marco Malusa.

JEL: C21, C53, D22, G30, J16, K22

Keywords: Generalized Random Forest, Causal Machine Learning in Double and Triple Difference, Covid-19, large firms, women on corporate boards, female leadership, quota

1 Introduction

During the Covid-19 pandemic, the notion of a superior crisis resilience of countries headed by female leaders further accelerated ongoing debates around female leadership in business and politics (Coscieme et al. (2020)). A dynamic public debate and research agenda have gained traction from a crisis resilience, gender equality and corporate governance perspective as part of rising ESG-investments[1]. However, with little truly exogenous appointment of women into leadership roles, it is difficult to identify the effect attributable to the gender (and correlated characteristics) of a leader. With some exceptions (e.g., Beaman et al. (2012)), the literature around female leadership, especially in the context of women on corporate boards, is reliant on quasi-experimental, fixed effects methods with limited grasp of complex, non-linear relationships.[2] In contrast, new Machine Learning methods bear the potential to detect more granular patterns of heterogeneity. Using a combination of Triple Differences and Causal Forests, this piece of research illustrates the value and limitations of the Causal Forest methodology when aiming to identify drivers of differences by the gender of leaders. The contribution thus lies both in the novel application of the Causal Forest method to a female leadership question and in the findings for the case of crisis resilience between Indian firms with and without female board members.

Method-wise, I build upon recent advancements in applying Machine Learning to causal questions. Topic-wise, I build upon recent research on the role of female leaders and gender quotas in (Indian) politics and business, with a focus on financial performance during time of crisis. While the literature largely agrees on positive effects of female politicians - mainly on creating public goods and igniting aspirations, findings in the business spectrum are rather heterogeneous (Ford et al. (2012), Sarkar and Selarka (2021), Singh et al. (2019)). By carefully exploiting two (to a certain

[1]Investments based on environmental, social and corporate governance criteria

[2]The complex corporate governance process might allow for effects of women on board only after initial acclimatisation time, to then accelerate non-linearly.

degree) exogenous events - the requirement for one woman on the board of large corporations as part of the Indian Companies Act 2013 and the 2020 economic shock of Covid-19 - this piece of research investigates the intersection of the gender quota and crisis resilience literature in gender development economics. To the best of the author's knowledge, there has not yet been research on the relationship between board room quotas and crisis resilience in an emerging economy, nor the application of Causal Forests in the broader female leadership context, as called for by Lee et al. (2020).

An initial Double Differences analysis finds that Indian firms complying with the 2013 requirement for one female director experience lower revenue declines during the economic shock of Covid-19 in 2020Q2/Q3, however, within the sectors most affected by Covid-19, the relationship between female directors and revenue changes is much weaker (Triple-Difference Analysis). While firm and time fixed effects are used to reduce endogeneity, compliance with the female director policy still is likely confounded by unobserved (non-constant) characteristics determining self-selection of women into boards of specific firms.

A subsequent novel application of Causal Forests in the firms literature enables first, to better cope with the issue of self-selection through forest-driven matching, and second, to enable a granular evaluation of the firm characteristics associated with the estimated effect. Using the advances in Causal Machine Learning for the latter has three advantages, (a) Using the richness of the data, so that dimensionality and multiple-hypothesis test restrictions of traditional methods can be overcome to (b) deeply analyse multi-dimensional relationships of the more than 50 characteristics and (c) find non-linear patterns in the unstructured inter-temporal evolution of company characteristics related to the prevalence of female leaders. I find that capital dynamics, size and the independence of female directors largely differ between firms with seemingly strong vs. weak effects of female leadership during Covid-19. Throughout the analysis, I directly employ robustness checks for each method, to enable a robust build-up of these findings.

The implications of this analysis are limited to 1) female leadership in large Indian companies[3]; 2) causal interpretations only under the strong assumption of no selection effects; so that the main contribution is in providing 3) suggestive evidence of characteristics correlated with the resilience of large female-led firms during Covid-19 in 2020, that illustrates the potential in combining Diff-in-Diff with the Causal Forest methodology for female leadership questions.

2 Methodology and data

2.1 Background

Historically, Gender equality has been neglected in economic development, especially regarding leadership positions (Sen (1999), Morrison et al. (2007), Mikkola (2005)). Even though female[4] leaders have been found to be partly more effective than their male counterparts in driving economic outcomes, they still lack equal access to powerful decision-making positions both in politics and business.[5]

In politics, quotas and reservation systems were successful in increasing female representation, aspirations of women and the provision of public goods (see Ford et al. (2012) for an overview). A famous example is the Indian reservation system introduced in 1993, requiring at least one third of village councils to be reserved for a woman chief councillor. Evaluations of this natural random experiment by Beaman (2010), Beaman et al. (2012) and Chattopadhyay and Duflo (2004) reveal

[3] Small and medium sized business are very differently affected during Covid-19, often female-led business can be worse off, see Liu et al. (2021))

[4] "Female" is used interchangeably with "non-male" in this article, being aware of non-binary genders, while data and literature constraints only allow for a simplified binary analysis

[5] Only 17% of board seats in the top 500 listed companies in India were held by women in 2020 (Dalal (2020)), 16.9% in the biggest 8600 companies worldwide in 2018 (Deloitte (2018)).

increased provision of public goods, reduced bribery, and higher educational attainment of girls, though cannot find improvements in women's labour market outcomes. According to these evaluations, the effects can be attributed to women as role models, as promoters of female preferences for public goods and potentially as less short-sighted than male leaders.

In business, initial research found positive (Farrell and Hersch (2005), Campbell and Mínguez-Vera (2008)), neutral (Rose (2007), Francoeur et al. (2008)) and negative (Wang and Kelan (2013)) relationships between board gender diversity and financial performance, based Difference-in-Differences and Fixed effects analysis. With the introduction of gender quotas for firm boards in the past 20 years, most recent literature still uses Double and Triple Differences, now embedded in policy and event studies with strong parallel trend assumptions of early and late adopters. Reported quota-impacts are positive on female representation beyond mandatory numbers, though evidence on business performance is mixed (Reddy and Jadhav (2019) for an overview[6]). In India, the 2013 Companies Act required all listed firms, and those exceeding a turnover of INR 300cr (approx. $ 40m) or paid-up share capital of INR 100cr Rupees, to appoint at least one female director from FY 2014/15 onwards[7]. Using a small panel of 20 firms, Srivastava et al. (2018) find that a higher number of female board members in Indian listed firms is associated with a higher return on assets. With a larger, but unbalanced sample of 1348 firms, Sarkar and Selarka (2021) conclude that listed firms with more female directors had a higher return on assets 2005-2014, and firms complying with the one-female-director-rule either in anticipation or when it became mandatory had a higher market value as measured by Tobin's Q[8]. They find that non-independence and close family ties diminish this effect. However, Sarkar and Selarka (2021)'s

[6] Additionally, Wang and Kelan (2013) (Norway), Rebérioux and Roudaut (2019) (France), Gutiérrez-Fernández and Fernández-Torres (2020) (Spain)

[7] By the end of 2021, there will also be crisis data available (annual reports FY2020/21) of firms that are just below this requirement for interesting discontinuity designs, but currently data are limited to quarterly reports of the largest firms that are all above this threshold.

[8] Tobin's Q equals the market value of a company divided by its assets' replacement cost

Difference-in-Differences analysis of the 2013 policy is prone to two types of endogeneity: First, unobserved firm characteristics determine both the performance and the (self-)selection of female directors into compliant and non-compliant firms - Sarkar and Selarka (2021)'s industry and time fixed effects only capture linear, constant, industry-level confounding. Second, reverse causality - when performing better, firms might be more likely to comply with the policy. I aim to handle both issues in my analysis that builds on the same policy; with a constant gender share, crisis addition, and a less constrained Machine Learning methodology.

Mechanisms for differential business performance associated with the gender of the leader are split between pre-appointment dynamics - the supply and demand of female leaders - and post-appointment leadership effectiveness. Lacking aspiration, aversion to competition and high costs of entry limit the supply of female managers; while taste, statistical and selection-system discrimination reduce the demand (Ford et al. (2012)). Once appointed, resource dependency theory suggests that female leaders might have access to different connections, information and leadership tools, that are more beneficial in reducing risks than the resources of another male leader (Hillman et al. (2009)). Agency theory adds that female leaders might exhibit more long-term thinking, better aligned to shareholders, as indicated by Nguyen et al. (2015).

During a crisis, the effect of female leadership might be more visible. First indications of a higher resilience of female-led companies come from the 2008 financial crisis - Palvia et al. (2015) find that female-led US banks were less likely to fail. Mechanisms could be a higher creation of social capital and more long-term planning that pay off during a crisis (Torres et al. (2019)), higher risk-aversion (Parrotta and Smith (2013)) and less overconfidence (Chen et al. (2019)).

With these potential channels in mind, the following analysis builds upon the quota literature (esp. Sarkar and Selarka (2021)) in the intersection with the crisis literature to understand the differential resilience of Indian firms during Covid-19. Data availability does not allow for an understanding of the leadership style or resources of female leaders, but rather of the characteristics of firms and firm boards into which they (self-)select or perform superior.

2.2 Identification strategy

To disentangle the effects and correlated factors of female leadership in Indian firms, the analysis builds upon a carefully identified timeline of shocks that requires balanced panel data. The 2013 Indian Companies Act mandated one female director in large or listed companies. It was likely anticipated with the first publicly accessible draft published in late 2011. After the anticipation period, April 1st, 2014 marks the date when one female director became mandatory through the Companies Act 2013[9] Firms could respond to the regulation in four ways: First, never-taking, by simply keeping the share of female directors at 0. Second, always-taking, as some firms already had one woman director beforehand. Third, just-complying, by adding one female director either in the anticipation period or in 2014. Fourth, defying, by reducing the number of female directors to zero due to internal company dynamics. Henceforth, always-taking and just-complying are called "complying", defying and never-taking are called "non-complying". Having none or 1+ women on board at the start of 2015 is exogenous under the assumption that compliance status is uncorrelated with any other firm or director characteristics that influence crisis resilience. This assumption is unlikely to hold, as pre-2011 take-up is likely correlated with firm characteristics. Omitting always-takers (and defiers) removes some degree of endogeneity, but the comparison between never-takers and just-compliers still builds on the assumption that more resilient firms do

[9] The Companies Act was again enforced through a similar policy by Regulation 17 of the Securities and Exchange Board of India which became mandatory at the start of 2015.

not have higher capacity and better connections to attract suitable female directors to comply with the policy. I thus stick with comparing never-takers to all others for most parts of the analysis.

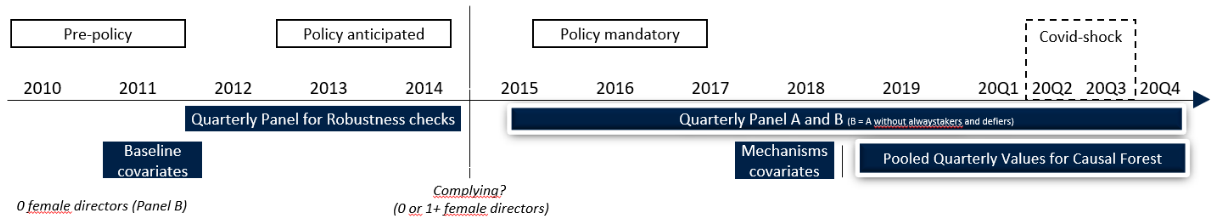

Figure 1: Timeline of events and panel setup
To have enough overlap, the panel needs to be restricted for the Causal Forest analysis (see section below). Baseline covariates are constant as of 2011/12, mechanism covariates are the growth of baseline covariates until 2018.

2.3 Firm panel

The Center for Monitoring the Indian Economy (CMIE) provides quarterly financial data and annual board member names and characteristics for the large Indian firms. From all 4120 Indian firms that published 2020 quarterly reports[10], 2783 have observations for each quarter from 2012 onwards, with details of annual board compositions of 2010-2020 and pre-policy, 2011/12 values of covariates. After rigorous data cleaning including winsorising the top and bottom 1% of covariate values[11], omitting any firms with missing values and further prediction-bias adjustments as described in the following subsection, this leaves a balanced panel of 2415 firms ("Panel A"). With this cleaning procedure, assuming that missing data is random, the subject of the analysis are "large Indian companies established before 2010".

To avoid reverse causalities, the main analysis only uses a constant value of the gender variable reported as of the start of 2015, the first measure after the Companies Act was enacted. Consequently, the quarterly firm panel is split into a panel for robustness checks covering the years 2012-2014, and the analysis panel that starts only after the constant gender measure, covering 2015Q1-2020Q4. Defiers, Always-takers, Just-compliers and Never-takers make up 2415 firms,

[10] All listed companies and the largest unlisted companies
[11] to omit special cases that are not the focus of this analysis of the "mainstream"

i.e. the 57960 firm-quarter observations in Panel A. Following the argument on endogenous self-selection above, I create Panel B that only covers firms without any women on board pre-announcement of the reform at the start of the years 2010 and 2011, which leaves 1207 firms (28968 firm-quarter observations).

2.4 Gender prediction

Even though the name of every board member is available, the gender is not. I quantify and adjust for errors of predicting the gender using the firstnames.[12] Using a combination of matching[13], a Bayesian classifier and a 300-node Neural Network[14], high prediction accuracies can be reached. The latter algorithms reach 80% accuracy on the non-board test data. For predicting the gender of board members, I start with a simple matching of all names with certainty >90%, and for the remaining names add the gender on which two of the three methods agree. Even though this combination should lead to an overall high accuracy in the region of 90-95%, I cannot rule out a possible prediction-caused correlation between board size (and thus company size) and the number of women on board. Consequently, I employ three checks and correction mechanisms: First, the top 10% largest boards are dropped (all with more than nine directors), which is especially important when differentiating between boards with no vs. at least one woman (see variable definitions below).[15] As seen in figure 2 below, after this adjustment, there is no systematic

[12] In difference to Sarkar and Selarka (2021) who do not quantify nor adjust for errors of their gender prediction.

[13] With a dataset of 900,000 Indian names obtained from genderize.io including certainty estimates

[14] Trained on 14,000 Indian names obstained from public records via GitHub, with inspirations from Kaggle(1) and Kaggle(2)

[15] Given that for each individual male and female names to be correctly predicted, the accuracy is at around 92.5%, then an all-male board of nine members would have a $0.925^9 \approx 50\%$ probability of being perfectly predicted as all-male. On the other hand, mispredicting boards that have one female member as zero females is unlikely ($0.075 * 0.925^8 \approx 4\%$ and up to 7.5% for smaller boards). Consequently, the effect of female directors might be overestimated if firms with higher boards sizes are systematically different from firms with lower board sizes (within the same sector for the Triple Differences later on). Although firms that are classified as compliers have on average boards with $+1.2$ ($se \approx 0.1$ for Panel A; for Panel B: 0.13 with $se \approx 0.13$) members in 2015, this mainly reflects that those firms simply increase their board size to include female director(s) (as seen by time trends: 2011 board size difference is 0.3 ($se \approx 0.12$ for Panel A; for Panel B: 0.15 with $se \approx 0.13$)).

relationship between compliance status and company size.

Nevertheless, the second correction mechanism is to rerun the analysis for a fixed board size, which leads to similar effect directions[16]. The third check of the prediction adds another layer of robustness: I use a totally different data source, Reuters Eikon, that provides pre-baseline values of the actual share of women directors for 35 very large Indian firms. As seen in the table, the prediction accuracy is reasonably high, $\frac{16+14}{35} = 86\%$ for the final gender variable. [17]

Table 1: Actual vs. Prediction of female directors variable

	Prediction (with Prowess names)		
Actual (Eikon Reuters)	No female directors	1+ female directors	Total
No female directors	16	3	19
1+ female directors	2	14	16
Total	18	17	35

35 firms of Panel A, 2012

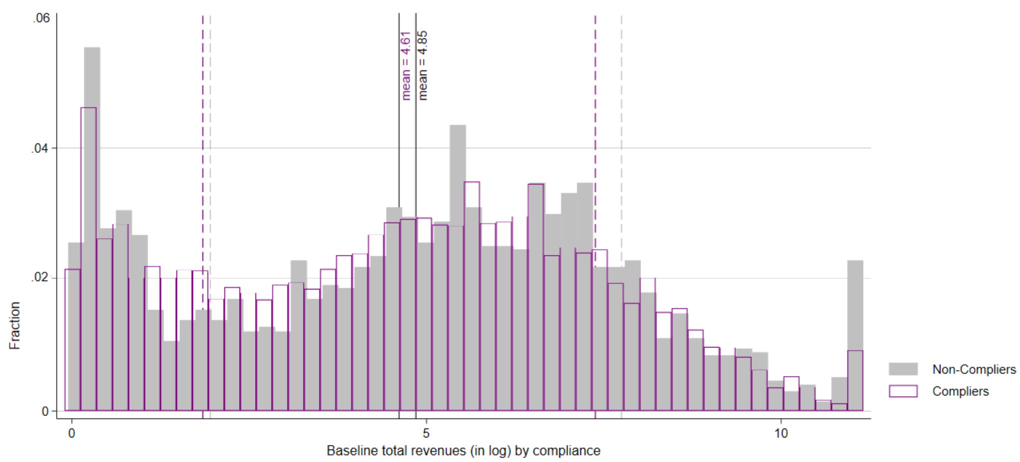

Figure 2: Histogram of company size in 2014
Measured in log revenues, with mean and standard deviation for compliers and non-compliers. This is for Panel A; Panel B similar.

[16] results available upon request

[17] The Reuters Eikon definition of "Board of Directors" is wider than prowess (includes CSR boards), leading to significantly higher board sizes and shares of female directors when these new type of boards were more common as part of board size expansions 2013 onwards - thus I use 2012 values. Additionally, the 35 very large firms available through Eikon Reuters might not be representative of the average firm in Panel A and B

2.5 Covid-19 shock

For quantifying the magnitude of the economic shock of Covid-19 by sector, I compare the sector-wise GDP of 2020Q2 with 2019Q2, as reported by the Indian Ministry of Statistics and Programme Implementation as of April 2021 (MOSPI). The final shock for each sector g is then $1.07 - \frac{GDP_{t=2020Q2}}{GDP_{t=2019Q2}}$[18]. Even though these values are derived from a different dataset, they could still be endogenous to female leadership: Studied (large) companies in a specific sector likely have more women on their boards historically when the sector is generally female-driven across company sizes. Thus, this female prevalence could have already reduced (increased) the economic shock by Covid-19 within a sector, so that I underestimate (overestimate) the total female contribution.[19] Based on overall GDP, country-wide mobility and trade data, I define the period in which the Covid-19 shock hit as 2020Q2 and 2020Q3 ("during Covid-19"), as the Indian economy was less affected in 2020Q1 and mostly recovered in 2020Q4. Any shocks in 2021 are beyond the data availability.

2.6 Variables

2.6.1 Firm resilience measure

With the target of understanding gender-induced heterogeneities in firm resilience, the literature suggests three different outcome variables: Tobin's Q, return on assets and total revenues (e.g. Park et al. (2009) and Vieira (2017)). As a side-effect of using recent data, asset values are not available in sufficient coverage, so that neither the return on assets nor Tobin's Q can be calculated; but the similarly often used, and for crisis evaluations suitable (change in) total revenues

[18] 1.07 represents the 7% average growth rate across sectors 2011-2019

[19] Other data sources to quantify the Covid-19-shock, like regional mobility and regional health outcomes on the demand and employee supply side, or trade and raw material supply data are either too granular for the nationwide activities of most companies in the sample, or incompletely depicting the combined shock.

can be obtained.[20] Thus, the outcome variable is defined as the logarithm of the total revenue of a quarter compared to the same quarter of the last year, i.e. for time (quarter-year combination) t and company i: $\Delta Y_{it} = log(Y_{it}/Y_{it-4})$.

2.6.2 Covariates

Covariates for controlling and explaining heterogeneities are twofold: First, characteristics of the firm (productivity, profitability, size, age, ownership) and second, characteristics of the directors (independence, executiveness, closeness to family) and board of the firm (share of directors by status, time of compliance).[21] To prevent reverse causalities, all covariates determining (self-)selection are held constant at the baseline, i.e. at pre-policy levels. Besides, the change of the same covariates from pre-policy to 2018 are captured as potential "mechanisms".

When aiming to estimate both, the average effect size, and quantifying heterogeneities, there is an ubiquitous trade-off between using covariates as controls or keeping them for explaining heterogeneities. In the Double and Triple Differences specification procedure, I use time and firm fixed effects to estimate the average effect size, keeping all covariates for explaining heterogeneities.

2.6.3 Female directors and shock variables

There are two sorts of "treatments": The gender treatment and the Covid-19 shock - the empirical hypothesis section explains when I use which. I follow the nature of the 2013 Indian Companies Act and simply differentiate between companies with zero and one or more female directors. To

[20]With the fixed effects setup later on, this outcome variable is further demeaned. If I would use not the change but the total outcome and then see the change via the fixed effects (FE) demeaning, there would be imprecision as this simple demeaning is more dependent on the number of periods covered than the comparison to previous quarters with an added FE demeaning of these comparisons across years

[21]More covariates like cash, loans and CEO remuneration are only available for an insufficiently small number of firms

prevent reverse causalities and measure at the point in time when the presence of female directors is most exogenously influenced (by the policy), the binary W_i treatment indicator is held constant at the level of the start of 2015. Whenever I refer to "firms with female directors on board" or "complying", I refer to this status as of the start of 2015.

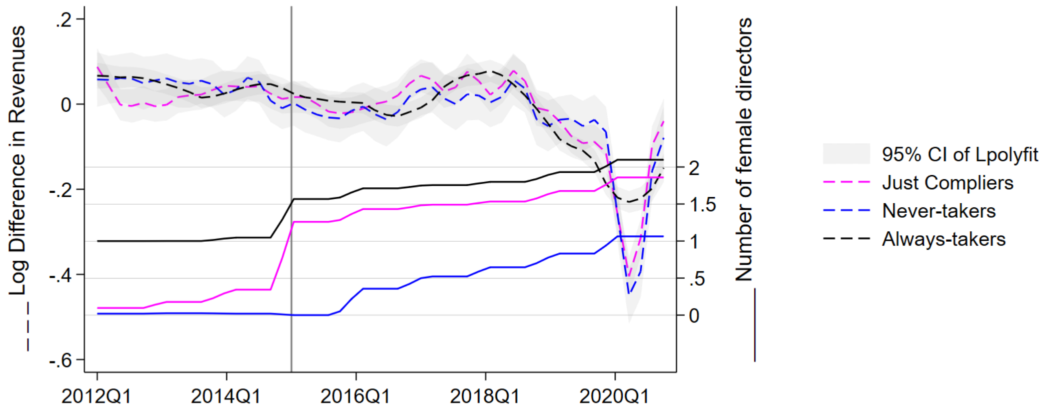

Figure 3: Outcome variable and number of women by compliance group pre- and post-policy
There are no significant differences before and after the policy between the three groups until Covid-19. During Covid-19, always-takers seem considerably better off, and just-compliers slightly better. Once adjusted for firm fixed effects, the difference between never-takers and just-compliers is more distinguishable, as seen in the regression outputs. Defiers are small in numbers and omitted here

3 Empirical analysis

3.1 Theoretical Hypothesis

Grounded in the reviewed female leadership and crisis resilience literature, the following theoretical model is assumed: Based on firm characteristics, female leaders are selected and self-select into sectors, firms and their boards. With the pressure of the policy, this process is accelerated. The presence of female directors then potentially affects financial performance, via the channels suggested by resource dependency and agency theory. However, a superior performance in crisis times cannot necessarily be attributed to women on the board, but rather be a consequence of the (self-)selection. Separating the two is the core challenge.

Even though there do not seem to be baseline or trend differences of the outcome variable between compliers, just compliers and always-takers pre-policy (see figure 3), there are also no differences visible from 2015 to 2019 in non-crisis times. If the differences are only visible during a shock[22], one might argue that the only way to prove baseline-balance would be a shock similar to Covid, pre-policy. I use a subset of the data and see no differences during the financial crisis 2008. However, the general shock of the financial crisis was very small on Indian companies, and secondly, its nature was very different. Thus, when the potential effect is only visible during a crisis, it is difficult to prove that non-complying firms are a valid counterfactual to compliers. The initially used Double Differences are based on this assumption. With Triple Differences, I improve this setup: I compare firms within a level of Covid-affectedness, thus the counterfactual is non-compliers who were in a similarly affected sector. This more granular definition is still dependent on a parallel trends assumption, though by restricting to within-sector comparisons, the assumption is more likely to hold.

3.2 Double Differences

In the first setup, I use Double Differences: The "treatment" is complying or not complying with the policy (i.e. having or not having female board members at the start of 2015). This treatment only comes to light during crisis times, so that effectively non-treated periods are all quarters except for 2020Q2 and 2020Q3. I estimate the average relationship of having at least one female director at the start of 2015 (W_i) and the change in revenues (ΔY_{it}), during quarter-year t with the following Double Differences equation for firm i with time and firm fixed effects[23] 1_t and 1_i:

$$\Delta Y_{it} = \alpha + \delta(1_{t=2020Q2|Q3} * W_i) + 1_t + 1_i + \epsilon \tag{1}$$

[22] "Resilience structures" that would make differences visible during non-crisis times are unobserved

[23] The firm fixed effect set-up in combination with a constant gender variable is more robust than using industry fixed effects or a lagged gender variable, but only captures the difference of having women on board in crisis times compared to non-crisis times, not a general effect of the female Dummy without time-interaction.

Hypothesis 1: $\delta = 0$. During Covid-19, firms that had female directors in 2015 have the same revenue growth as firms without.

Table 2: Double Differences Fixed Effects Regression

	(1) Panel A	(2) Panel B
2020Q1	-0.178***	-0.159***
	(0.0240)	(0.0340)
2020Q2	-0.599***	-0.596***
	(0.0351)	(0.0444)
2020Q3	-0.254***	-0.231***
	(0.0338)	(0.0413)
2020Q4	-0.0596**	-0.0430
	(0.0231)	(0.0327)
2020Q2/Q3=1 $\times$ W=1	0.0712**	0.0668*
	(0.0315)	(0.0391)
Constant	0.0305*	0.0159
	(0.0165)	(0.0228)
Observations	57960	28968

Robust standard errors in parantheses. Further Firm and Time FE included. *** p<0.01; ** p<0.05; * p<0.1

The Double Differences regression rejects Hypothesis 1 - firms with female directors in 2015 experienced a significantly smaller loss in revenues during Covid-19. Compared to the growth of non-complying firms, they grew $exp(0.0712) - 1 = 7.4\%$ more in Q2/Q3, c.p. As a robustness check, when excluding all always-takers and defiers, the association stays at a similar level (6.9%, Panel B).[24]

The counterfactual used here is the trend in non-compliers, $\Delta Y^{W_0}_{i,t=2020Q2|Q3} - \Delta Y^{W_0}_{i,t\neq 2020Q2|Q3}$. A causal interpretation depends on the exogeneity assumption that without female directors, the compliers would have had the same trend in ΔY_{it} as the non-compliers. When female leaders simply self-select into sectors less affected by Covid-19, like Professional Services, this assumption is violated. With the following more granular Triple Difference design, exogeneity only needs to hold within a sector.

[24] A Placebo Hypothesis of defining the crisis for 2017Q2/Q3 yields no statistically significant association of female directors with revenue change

3.3 Triple Differences

Table 3: Sector wise shock

	Panel A			Panel B		
	Firms	Compliers	SectorShock	Firms	Compliers	SectorShock
Agriculture	19	.74	.04	10	.5	.04
Mining	26	.73	.30	12	.75	.30
Manufacturing	1045	.79	.46	559	.70	.46
Utilities	13	.77	.14	7	.71	.14
Construction	119	.80	.57	67	.76	.57
Trade, Transport and Services	523	.83	.54	250	.80	.54
Professional Services	587	.83	.12	268	.80	.12
Public Sector	83	.81	.17	34	.59	.17
Total/Average	2415	.78	.30	1207	.70	.30

Overview of the sector-wise shock specifications in Panel A and B, calculation as described in Section 3.4.

For Triple Differences, I add the new Covid-19-shock C that differs by sector g to the Double Differences equation 1. The previous table provides a sector-wise magnitude of this shock. With firm and time fixed effects, only the interaction of Covid-time with each of the other two remains.

$$\Delta Y_{it} = \alpha + \beta(1_{t=2020Q2|Q3} * W_i * C_g) + \delta^{tri}(1_{t=2020Q2|Q3} * W_i) + \eta(1_{t=2020Q2|Q3} * C_g) + 1_t + 1_i + \varepsilon_{it} \quad (2)$$

The Triple-Differences design could be understood as a reframing of the "treatment" as the intensity of the Covid-19 shock, with the counterfactual being firms that operate in less-affected sectors. Then, the core interpretation is whether, within strongly affected sectors, firms with female directors on board (as of 2015) exhibit lower or higher financial resilience. For inequality assessments, a descriptive analysis like this is relevant as such; for inferring causality the following assumption is needed: A valid counterfactual for complying firms are non-complying firms within the same level of sector-wise Covid-affectedness. This is a much weaker assumption than for Double Differences, because, as the literature suggests (e.g., Terjesen et al. (2016)), the main (self-)selection bias lies in sectoral allocation.

Hypothesis 2a: $\beta = 0$. During Covid-19, firms with female directors in strongly affected sectors have the same revenue growth as firms without female directors or firms in not affected sectors.

Hypothesis 2b: $W_i\beta + \eta = 0$. During Covid-19, within each level of compliance, firms in strongly affected sectors have the same revenue growth as firms in not affected sectors

Hypothesis 2c: $C_g\beta + \delta^{tri} = 0$. During Covid-19, within each level of sector-affectedness, firms that had female directors in 2015 have the same revenue growth as firms without.

Table 4: Triple Differences Fixed Effects Regression

	(1) Panel A	(2) Panel B
2020Q2/Q3 × Covid-Sector-Shock	-0.656***	-0.630***
	(0.182)	(0.228)
2020Q2/Q3 × W	0.296***	0.328**
	(0.0886)	(0.128)
2020Q2/Q3 × W × Covid-Sector-Shock	-0.595***	-0.671**
	(0.217)	(0.294)
Constant	-0.0190***	-0.0161**
	(0.00609)	(0.00810)
Observations	57960	28968

Robust standard errors clustered by firm in parantheses. Firm and Time FE included. *** p<0.01; ** p<0.05; * p<0.1

When interacting female leadership with a continuous sector-wise shock, the estimate is negative, Hypothesis 2a is rejected. Similarily, Hypothesis 2b is rejected[25]. To illustrate this, a typical difference in sector-wise shock, for example between manufacturing and professional services is $0.46 - 0.12 = 0.34$ (see table 3). At $W_i = 1$ (compliance) and $1_{t=2020Q2|Q3} = 1$, a stronger sector-shock by 0.34 would lead to a growth difference of $exp(-(0.595 + 0.656) * 0.34) - 1 = -35\%$. A firm in the manufacturing sector compared to the professional services sector thus experiences a stronger shock, and, this shock is worse when firms have female directors ($-35\%\ vs. -20\%$).

[25] F-test statistic is 112.85, given W=1

Hypothesis 2c cannot necessarily be rejected - it needs to be put in the context of levels of the sector-shock: The derivative of equation 2 by W_i (with $1_{t=2020Q2|Q3} = 1$) yields $\beta * C_g + \delta^{tri} = -0.595 * C_g + 0.296$. Holding the Covid-Sector-Shock C_g fixed at 0, the positive association between crisis resilience and female leaders seen in the Double Differences is confirmed, for all firms with $C_g < 0.4974$ (or $C_g < 0.255$(**)[26] including 95% confidence intervals) - i.e. for all firms in Agriculture(**), Mining, Manufacturing, Utilities(**), Professional Services(**) and Public Sector(**), the presence of female leaders given a Covid-19 shock in the sector is positive (compare with table 3). In sectors stronger affected by Covid-19, this positive association dissipates. Once always-takers and defiers are excluded, the results do not change significantly. Further robustness checks, including a differently calculated sector-shock[27] lead to similar results, a placebo crisis for 2012 and 2017 does not yield statistically significant results[28].

Both ways of interpretation are interesting - either evaluating the association of the shock and financial resilience by the prevalence of female directors (non-causal female leadership interpretation; Hyp. 2b), or evaluating the association between female directors and financial resilience by different shock levels (causal female leadership interpretation; Hyp. 2c). Both have the most interesting part of the estimation in common: the interaction term β. β indicates the association of female directors in strongly affected sectors with financial performance and thus is of especially high interest. With an understanding of the heterogeneities of this association by firm and director characteristics I can disentangle underlying potential drivers of female-led firms' crisis resilience. For this analysis, I need to move beyond Triple Differences and traditional econometrics. Triple Differences only capture a more precise Covid-19 shock by sector-affectedness, but including further interactions of this Triple Difference term with firm characteristics is infeasible as their number

[26] $0.255 = -(0.296 - 0.089)/(-0.595 - 0.217)$

[27] Using 2020Q3 shock values or a random change of +/- 0.1 of C_g to simulate measurement inaccuracies

[28] Available upon request

would become prohibitively large and estimates fuzzy. To still understand heterogeneities, I instead use a Machine Learning based partitioning and estimation technique called "Causal Forests".

3.4 Causal Forest

3.4.1 Theoretical background

Causal forests (Wager and Athey (2018)) as a special case of generalised random forests (GRF, Athey et al. (2019)) is a method of non-parametric statistical estimation of heterogeneous treatment effects. As an application of random forests (Breiman (2001)) for causal inference, the method builds on the potential outcome model of Rubin (1974) and the classical set of conditional unconfoundness assumptions, 1) conditional independence, 2) stable unit treatment value assumption, 3) overlap and 4) exogeneity of covariates (Lechner (2019), Imbens (2000)). The goal of applying Causal Forests is to identify heterogeneities in treatment effects, without the problems of traditional econometrics (of imposing assumptions on the functional form and dimensionality) and without the limitations of predictive Machine Learning (of purely correlational analysis and no valid confidence intervals).

Causal trees are the foundation. A tree partitions the data into subsamples by minimising the expected means squared error (EMSE) between true $\tau(x)$ and estimated treatment effects $\hat{\tau}(x)$ through $max\ [E(\tau^2(x)) - E(Var(\hat{\tau}(x)))]$, thus growing a tree by answering *"Where can we make a split that will produce the biggest difference in treatment effects across leaves, but still give us an accurate estimate of the treatment effect?"*. After growing the tree, an average treatment effect per subgroup is estimated. Athey and Imbens (2016) show that, given a vector of covariates x, τ is not needed for the estimation of $\widehat{EMSE}$. After splitting the data into a first sample S^{tr} with N^{tr} observations for growing the tree and a second sample S^{est} with N^{est} observations for estimating the treatment effects per leaf l (including confidence intervals based on the within-leaf variance $L^2(l)$ for a partition Π of the covariate space), a decision tree is grown using the expected mean squared error of the estimated treatment effect

$$\widehat{EMSE}_{\tau}(S^{tr}, N^{est}, \Pi) = -\frac{1}{N^{tr}} \sum_{i \in S^{tr}} \hat{\tau}^2(X_i; S^{tr}, \Pi) + (\frac{1}{N^{tr}} + \frac{1}{N^{est}}) \sum_{l \in \Pi} (L^2_{S^{tr}}(l(x; \Pi))) \tag{3}$$

plus tuning parameter Φ as the splitting criterion: $argmin\ \widehat{EMSE}_{\tau}(S^{tr}, N^{est}, \Pi) + \Phi(no.\ of\ splits)$, with Φ choosen by cross-validation. As seen in the equation, the minimisation of the EMSE is a trade-off between the variance of treatment effects across leaves (first term) and the uncertainty about leaf treatment effects (second term). Finding more heterogeneities is, after the optimal point, only possible by growing leaves that yield very imprecisely estimated effects.

The balance between the two in growing the tree can be entirely achieved with the first part of the data S^{tr}. Then, the second part of the data is used to estimate the effects. This split and its resulting exclusion of Y^{est} (the outcome used for estimation) from growing the tree is termed "honesty", which allows for an unbiased estimate under the unconfoundedness assumptions Athey and Imbens (2016). In observational studies, causal trees are usually build with the treatment propensity to reach conditional unconfoundedness. These propensity scores are estimated via "propensity trees" - the prediction of the treatment using covariates. This approach is similar to other propensity score methods in their way of reaching conditional unconfoundedness by quasi-matching on similarity in covariates X. However, as Chernozhukov et al. (2018) (based on Funk et al. (2011) and Kennedy et al. (2017)) show, these propensity trees yield a doubly robust estimator, i.e. they are unbiased even if only one of the two models - the propensity score model or the outcome model - is correctly specified (see Appendix).

While one Causal Tree deals well with high dimensionality and functional form assumptions, a forest of trees is more precise (Wager and Athey (2018)). Currently[29], the best way to aggregate trees to a forest is through decorrelation: Only a random part of the sample is used to grow every tree and only a random subset of all covariates can be used for any split. Then, the forest outcome is calculated via a nearest neighbourhood weighted average of predicted treatment estimates across trees (Breiman (2001), Athey et al. (2019)). Wager and Athey (2018) prove that random forests and the estimated $\hat{\tau}(x)$ are asymptotically normally distributed and unbiased under the classical set of unconfoundness assumptions and additionally, honesty, uniform distribution, Lipschitz-continuity and reasonable scaling of covariates[30]. This is important as it allows for the calculation of valid confidence intervals via the variability of tree-specific treatment estimates, that use weighting to consider the random subsampling procedure. Specifically, Athey et al. (2019) and Efron (2014) show that under the mentioned assumptions, there is a sequence $\sigma_n(x)$ for which $\frac{\hat{\tau}_n(x)-\tau(x)}{\sigma_n(x)} \sim N(0,1)$, so that valid confidence intervals for $\hat{\tau}(x)$ can be constructed with $\sigma_n(x)$ asymptotically. $\hat{\sigma}^2_n(x)$ is estimated via a bootstrapping-inspired infinitesimal jackknife procedure of within-subsample half-sampling. With this, the between group and within group variation can be asymptotically correctly estimated when the number of trees B is sufficiently large.[31]

3.4.2 Applied procedure

The value of Causal Forests for the application to the female leadership and crisis resilience literature lies in the non-parametric detection of (non-linear) heterogeneities in a high-dimensionality context. To see under which firm and board characteristics firms with female directors in more strongly affected sectors exhibit higher financial resilience, I look at four core outputs of Causal Forests: (i) the re-estimation of the average β, (ii) subgroup-wise β when firms are split into

[29] The Machine Learning literature evolves fast - new forms like Boosted Regression Forest or Virtual Adverserial Training are promising, but are still considered experimental in the causal ML context, while Random Forests are established as a foundation of causal ML (Wager et al. (2020))

[30] More on these assumptions in the application below and in Wager and Athey (2018)

[31] For details, see Athey et al. (2019) especially

three subgroups along β, (iii) importance of firm and board characteristics[32] for growing the heterogeneity trees and (iv) the firm and board characteristics special to each of the three subgroups.

There is not yet an established procedure of how to apply Causal Forests in panel data. Pioneering researchers, e.g., Miller (2020) and Burauel and Schroeder (2020), have simply added data pre-processing steps and subsampling adjustments in their applications to convert a panel via fixed-effects residualisation into a pooled cross-section that can be used in a Causal Forest. Panel unit subsampling and standard errors clustered by firm are important to estimate valid confidence intervals, even after firm-wise demeaning (Miller (2020)). For firm i, unobserved constant linear effects are excluded, but all non-linear determinants that are shared for all t for a firm i might still downward bias the variance estimate. With panel subsampling, a tree is grown either with all or with no observations from a given panel unit. Orientated on Miller (2020) and Burauel and Schroeder (2020)'s suggested pre-processing and Athey et al. (2019)'s[33] R-learner-inspired general random forests estimations , I use the following nine-step procedure:

Step one: Obtain the residuals ε_{it}^{Y} from the Triple Differences equation 2 without the Triple Difference term WCT_{it}[34], $\Delta Y_{it} = \alpha + \delta^{Y} 1_{t=2020Q2|Q3} * W_i + \eta^{Y} 1_{t=2020Q2|Q3} * C_g + 1_t + 1_i + \varepsilon_{it}^{Y}$

Step two: Obtain the residuals ε_{it}^{WCT} from a regression of the Triple Differenced term on the remaining terms, $WCT_{it} = \alpha^{WCT} + \delta^{WCT} 1_{t=2020Q2|Q3} * W_i + \eta^{WCT} 1_{t=2020Q2|Q3} * C_g + 1_t + 1_i + \varepsilon_{it}^{WCT}$

Step three: Define $forest\ outcome\ variable\ \varepsilon_{it}^{Y} := \hat{\beta}^{forest}(\varepsilon_{it}^{WCT}) + \widehat{error}$

[32] There are 46 firm and board characteristics supplied to the forest, incl. mechanisms and baseline values of financial and non-financial variables. Detailed list available upon request

[33] Including Athey's GitHub code

[34] $WCT_{it} = W_i * C_g * 1_{t=2020Q2|Q3}$

Step four: Estimate the marginal expectations conditional on the constant, pre-policy covariates[35], i.e. the response functions $\hat{e}(x) = E(\varepsilon_{it}^{WCT}|X = x)$ and $\hat{m}(x) = E(\varepsilon_{it}^{Y}|X = x)$ using two separate regression forests with panel subsampling

Step five: Use the obtained nuisance components $\hat{e}(x)$ and $\hat{m}(x)$ to estimate $\hat{\beta}^{forest}$ for each point with, adaptive similarity weights $a_i(x)$ specific to each of the B trees in the random forest. [36] The similarity weights reflect the relevance of the i-th firm for the estimation of $\hat{\beta}^{forest}$ at test point x. Specifically, a_i represents the frequency of i-th firm in the same leaf L_b as x: $a_i(x) = \frac{1}{B}\sum_{b=1}^{B} a_{bi}(x)$ with $a_{bi}(x) = \frac{1(\{X_i \in L_b(x)\})}{|L_b(x)|}$ [37]

$$\hat{\beta}^{forest}(x) = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}[a_i(x)(\varepsilon_{it}^{Y} - \hat{m}^{(-i)}(X_i))(\varepsilon_{it}^{WCT} - \hat{e}^{(-i)}(X_i))]}{\sum_{i=1}^{n}\sum_{t=1}^{T}[a_i(x)(\varepsilon_{it}^{WCT} - \hat{e}^{(-i)}(X_i))^2]} \tag{4}$$

by using a regularisation (Θ)[38] and out-of-bag $^{(-i)}$ cross-fitting approach that minimises the

[35] This implies that every observation of the same firm will have the same propensity of treatment - thus panel subsampling is key to require outcome-comparisons between and not just within firms.

[36] Following suggestions by Athey et al. (2019) and Wager et al. (2020) for small samples, I use 70% of the data for building the tree and 30% for estimating heterogeneities ("honesty fraction") and use randomly chosen 50% of the observations for growing each tree. Given the imbalance of the sample I set a minimum boundary for the final node size of 10. Further, I use cross-validation to the remaining parameters, but cannot find an improvement upon the default of $\sqrt{number\ of\ covariates} + 20$ for the number of covariates randomly offered for each split, 5% for the maximum imbalance of a split and no further imbalance penalty. This setup should prevent leaves without sufficient observations, nevertheless I conduct sensitivity checks of the most arbitrarily chosen parameters final node size and maximum imbalance - see results below. The remaining parameter to be set is the number of trees B. To estimate the asymptotic variance of causal forest correctly with the jackknife procedure, the number of trees needs to be large enough that the Monte Carlo variability of the forest does not matter, which is the case when the by Athey et al. (2019) suggested rule of thumb $\#$ trees $\approx$ $\#$ individuals is fulfilled. With $n = 2415$ firms I set the number of trees to $B = 3000$. Tests with more trees do not significantly change the results in any way, suggesting that it is indeed large enough

[37] For simpler understanding, this depicts the binary treatment case. For continuous treatment the estimation is similar, though the variance in the denominator needs to be estimated via a mini-random forestAthey and Wager (2021). Specifically, the appendix provides a short overview of the average treatment effect when estimated with clusters and augmented inverse propensity weighting

[38] The regularisation is a result of the parameters described above, i.e. minimum node size, imbalance penalty and pruning.

squared loss function

$$\hat{\beta}^{forest}(\cdot) = argmin_{\beta^{forest}} \sum_{i=1}^{n} \sum_{t=1}^{T} [(\varepsilon_{it}^{Y} - \hat{m}^{(-i)}(X_i)) - \beta^{forest}(X_i)(\varepsilon_{it}^{WCT} - \hat{e}^{-i}(X_i))]^2 + \Theta(\beta^{forest}(\cdot)) \quad (5)$$

Step six: Empirically test the model calibration (see Hypothesis 3 and 4 below)

Step seven: Assess variable importance through ranking covariates by the number of splits for which they were used across trees

Step eight: Split the estimation space into three equally large subgroups along $\hat{\beta}^{forest}$

Step nine: Calculate subgroup-wise averages of covariates to see heterogeneities

3.4.3 Assumptions for application

Purged of time and firm fixed effects as well as Double Differences, using the residuals enables an exclusive focus on heterogeneity of the Triple-Difference term WCT and allows for pooling of the observations across quarters[39]. The average partial "effect" is then, more specifically, the "effect" of having female directors in 2015 within a level of the sector-wise Covid-19-shock on the log revenue difference during 2020Q2 and 2020Q3, striped off firm FEs, time FEs and the shock-time and female-time interactions.

[39] Striping off fixed effects, fulfils the main assumption needed for pooling: Afterwards, there are no correlations between 1_i, and the time-constant covariates. Besides, the applied linear cleaning of the the Double Differences might be insufficient, as β^{forest} and δ^{tri} could still be negatively correlated for subgroup-wise estimates. However, this negative correlation appears to be only small, so that interpretations assuming a constant δ^{tri} across subgroups split along β^{forest} are reasonable

For estimating an unbiased causal relationship and the terming "effect" to be appropriate, the following six main assumptions need to hold for the data used in the causal forest estimation. This data are Panel A (2019Q1-2020Q4, see explanation below) after pre-processing steps one to three, i.e. after purging.

1) Conditional Independence: The use of firm fixed effects to purge unobserved constant differences and propensity forests to de-confound the endogenous selection into complying and non-complying might reach close to fulfil conditional independence. Nevertheless, unobserved time-varying differences, for example the network of male directors for finding alternative buyers might mostly be relevant during a crisis, and if these male directors are then also better connected to recruit female directors, conditional independence is violated. Additionally, the intensity of the Covid-19 shock might be different between firms "treated" with female directors in a specific sector and those that were not. Even after adjusting for probably the largest confounding factor - sectoral differences - with the triple-difference approach, granular unobserved differences within sectors, like firm agility, government connections and international network might be correlated with the (self-)selection of female directors.

2) SUTVA: Even though there might be competition between the firms for qualified female directors, especially since most firms substantially increased their number of female directors at the same time in 2014, this competition is not judged as relevant enough because of the vast spread across industries and a constant inflow of highly-educated labour with strong (international) experience (Brown and Sung (2015)). Besides, there are dynamic mainstream effects that might pressurise non-compliers - by using a constant 2015 value of the female director variable, I deal with this issue.

3) Overlap: The counterfactual is a combination of observations of non-compliers throughout all periods and compliers in non-Covid-19-periods.[40] WCT>0 only in 2 out of 24 quarters of 2015-2020 - this leads to 1788 vs. 27180 observations with WCT=0. As the overlap is critical, and the Triple Difference estimates for Panel A compared to Panel B not substantially different (see table 2), I include the 1063 always-takers (and with them, the 157 defiers), reaching a total of 3914 observations with the Triple Difference Term >0 (vs. 54046 with $WCT = 0$). This is very unbalanced, so that I further cut out all pre-2019 observations after the pre-processing, to reach a more balanced setup of 3914 with $WCT > 0$ vs. 15406 with $WCT = 0$. This still makes relatively wide definitions of neighbourhoods for stable comparisons necessary. Thus, even with a higher number of trees, the results should be taken with care. On the first glance, one could critically raise that the neighbourhoods of a treatment estimate that are used as a counterfactual are simply pre-2020 observations from the same firm as they have the same treatment propensity because of constant covariates. However, a look at the propensity score suggests that there are reasonably many firms across propensity neighbourhoods (see overlap figure below, further issues discussed in Hyp. 3).

4) Exogeneity of covariates: By holding all firm characteristics constant to pre-policy levels, there is no possibility of a reverse effect of female leaders on firm characteristics. Then, $X_i^{WCT=0}$ is equal to $X_i^{WCT>0}$ if neither the time, nor the intensity of the Covid-19-shock were anticipated. However, any (self-)selection based on characteristics relevant to the shock violates this exogeneity criterium. Additionally, the covariates describing the compliance status and mechanisms are purely descriptive and cannot be conditioned on.

5) Honesty: I use each cross sectional unit i and its outcomes ε_{it} either to estimate the within-leaf

[40] To be precise, because of the continuous nature of WCT, all levels of WCT (which is used in the Causal Forest after being transformed to ε_{it}^{WCT}) besides the one of a specific observation are counterfactuals

treatment effect, or to decide where to place the split, not both.

6) Uniform distribution, Lipschitz continuity and appropriate scaling: Orientated on Wager et al. (2020), those three are fulfilled for the given setup: The number of observations is greater than the number of parameters n>p, the first derivatives are bounded, covariates scaled pre-analysis and the algorithm requiring a subsample size scaling appropriately[41] with n.

In summary, one should be sceptical about causal interpretations of the results, like for the Double and Triple Differences, because even after preprocessing and further deconfounding with treatment propensity, there likely are considerable unobserved factors that confound the estimated treatment effect. The little variation in the propensity score estimation illustrates this issue: To check the overlap and performance of the propensity $\hat{e}(x)$, I test the following two hypothesis:

Hypothesis 3: The predicted propensity score is significantly better suited than the naive propensity $\hat{e}^{naive}(x) = \frac{1}{n}\frac{1}{t}\sum_{i=1}^{n}\sum_{t=1}^{T}\varepsilon_{it}^{WCT}$. This is assessed by: 3a) $\hat{e}$ has sufficient overlap and 3b) $\hat{e}$ is significantly different to $\hat{e}^{naive}$

The figure suggests that observations with low and high levels of treatment intensity ε^{WCT} overlap sufficiently. A reason for this overlap is that $\hat{e}(x)$ is only spread in a very small part of the true ε^{WCT}, so that in fact, it is not much different from the naive $\hat{e}^{naive}(x)$, Hypothesis 3b cannot be rejected. This does not imply randomness but more likely that the observed characteristics and number of observations are insufficient for explaining the (self-)selection process, so that the results are roughly the same when using the naive propensity. [42]

[41] Appropriate scaling means that the subsample size s scales as $s \asymp n^{\psi}$ for some $\psi_{min} < \psi < 1$ (Wager and Athey (2018) and Wager et al. (2020))

[42] Even when restricting the sample to only 2020Q2 and 2020Q3 to avoid a downward reduction of the propensi-

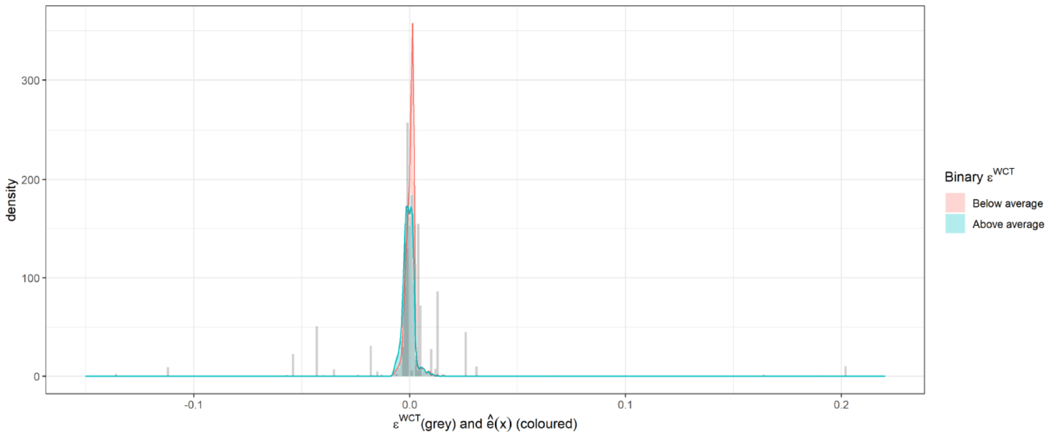

Figure 4: Density of $\hat{e}(x)$ and ε^{WCT}
$\hat{e}(x)$ is coloured by below or above-average ε^{WCT}, which shows good overlap but strong focus around 0. Density of $\hat{\beta}^{forest}$ in grey

Hypothesis 4: 4a) The average prediction produced by the forest is correct ($\upsilon = 1$) and 4b) the forest does not capture any heterogeneity ($\varphi = 0$)

$$\varepsilon_{it}^{Y} - \hat{m}^{-i}(X_i) = \upsilon \overline{\beta}^{forest}(\varepsilon_{it}^{WCT} - \hat{e}^{-i}(X_i)) + \varphi(\hat{\beta^{forest}}^{-i}(X_i) - \overline{\beta}^{forest})(\varepsilon_{it}^{WCT} - \hat{e}^{-i}(X_i)) + \epsilon \quad (6)$$

As the following table shows, the average prediction is close to 1, Hyp. 4a cannot be rejected. Hypothesis 4b is a popular "omnibus test for heterogeneity"(Chernozhukov et al. (2018)); φ measures how the predictions for the conditional APE covary with the true conditional APE. Here, Hypothesis 4b cannot be rejected, suggesting that there might not be strong heterogeneity detectable with the used covariates. This result of at most minor heterogeneities is confirmed in several robustness checks (see below). Thus, the following approach is based on a split into just three subgroups, that might nevertheless be able to capture some heterogeneity.

3.4.4 Results

(i) The original β of the Triple Differences is re-estimated with the Causal Forest method.

Hypothesis 5: $\beta^{forest} = \beta$ (The estimated Triple Differences coefficient is the same in a regression and in a doubly robust Causal Forest)

ties for firms with $WCT_{i,,t=2020Q2|Q3} > 0$ (because in pre-crisis periods, their $WCT_{i,t<2020Q2} = 0$), there remains this imprecision.

Table 5: Calibration check of Causal Forest

	$\varepsilon^Y_{it} - \hat{m}^{-i}(X_i)$:
υ	1.057***
	(0.386)
φ	0.406
	(0.555)

*p<0.1; **p<0.05; ***p<0.01; N=19320

With an estimated average partial "effect" (APE) of $\hat{\beta}^{forest} = -0.52(se : 0.49)$[43] estimated in the Triple Differences regression as shown in the figure below. Hypothesis 5 cannot be rejected. This is the case even though the Causal Forest has been further - weakly - deconfounded with doubly robust propensity scores. This confirmed avg. negative association between firms in highly affected sectors with female board members could be linked to resource dependency theory: As Danes et al. (2009) suggest, the networks of female directors might be worse than the one of male directors, so that female-led firms are less able to secure government contracts in a crisis. However, as I cannot separate the performance attributable to female directors from the (self-)selection this interpretation should be taken with care. Instead, a comparative look at the subgroup that exhibits a positive (negative) association might indicate ways (barriers) of (self-)selection and influence.

(ii) When splitting $\hat{\beta}^{forest}$ into three subgroups along its effect size, to test:

Hypothesis 6: $\hat{\beta}^{forest}_1 = \hat{\beta}^{forest}_2 = \hat{\beta}^{forest}_3$ (H: No difference between subgroup-wise APE)

the Causal Forest detects joint statistically significant differences at 10% (Wald-test p-value=0.085), driven by the difference between subgroup 1 and 3 (see figure above, p-value=0.03 for $\hat{\beta}^{forest}_1 = \hat{\beta}^{forest}_3$). This result suggests that even when general tests do not suggest strong heterogeneity (see Hypothesis 4b), specific analysis might still find some. To understand the drivers of this heterogeneity, there is high value in comparing which covariates differ the most between subgroup 1 and 3. For

[43] The standard error is higher due to reduced sample size from 57960 to 19320

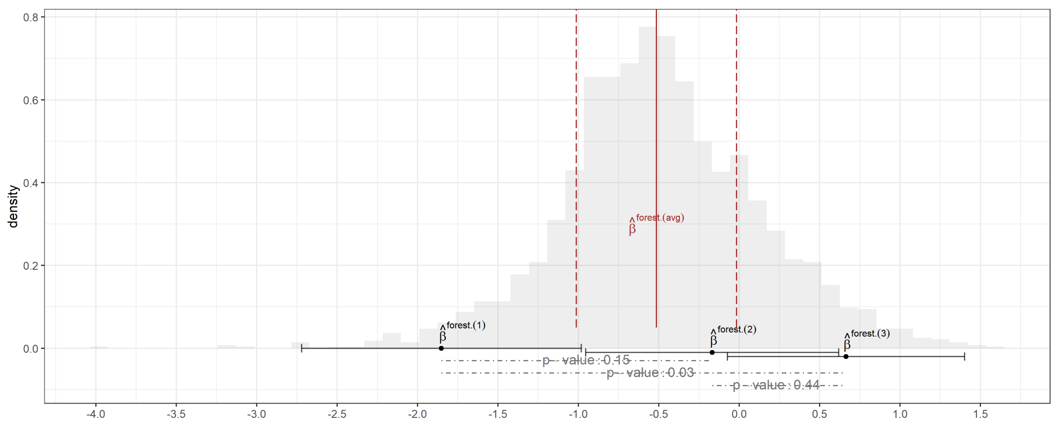

Figure 5: Distribution of $\hat{\beta}^{forest}$ and $\hat{\beta}^{forest}$ by subgroup

this, I first use the variable importance for growing the Causal Forests as an indicator in (iii), to then specifically look at covariate differences between subgroups in (iv).

(iii) The importance of each firm and board characteristics j for growing the forest is assessed with a standard counting of the number of times it was used for splitting in the first four layers k across all trees B (O'Neill and Weeks (2018),Wager et al. (2020)):

$$imp(x_j) = \frac{\sum_{k=1}^{4}\left[\frac{\sum_{b=1}^{B} number\ depth\ k\ splits\ on\ x_j}{\sum_{b=1}^{B} total\ number\ depth\ k\ splits}\right]k^{-2}}{\sum_{k=1}^{4} k^{-2}} \tag{7}$$

The covariate importance measure suggests that the share of independent directors in 2015, the labour productivity baseline 2011 and change 2018 vs. 2011 and the total wages at baseline and its change are most relevant for heterogeneities in the association of female leadership and the Covid-shock with revenues. These covariates could potentially drive crisis resilience. However, firms with higher productivity and more resilient business models, might generally attract more women as directors. The influence of them as a director could be comparatively small, and most of the heterogeneity explained by (self-)selection. The simple importance measure (7) should be taken with care. If two covariates are highly correlated, the trees might split on one covariate but not the other - thus covariates with seemingly low importance might in fact be highly relevant for heterogeneity, but simply correlated with another, slightly better predicting covariate.

(iv) The firm and board characteristics special to each of the three subgroups are illustrated in figure 6, by calculating where the mean value of a covariate in a subgroup lands on the standardised empirical distribution of this covariate. If the colouring differs between subgroups for a covariate, then the covariate is associated with the heterogeneity in $\hat{\beta}^{forest}$.

		Subgroup 1	**Subgroup 2**	**Subgroup 3**	Covariate mean	Covariate sd
Firm Characteristics	Labour Productivity (Baseline 2011)	1.86602	2.5135	2.56143	2.31365	1.14654
		0.01708	0.01313	0.01008	2.31365	1.14654
	Total Wages (Baseline 2011)	2.06111	2.7307	2.99297	2.59493	1.94337
		0.02311	0.0245	0.02352	2.59493	1.94337
	Profits (Baseline 2011)	42.2132	139.798	212.046	131.352	608.524
		4.65664	7.68593	9.46066	131.352	608.524
	Change in total wages (Mechanism 2018 vs. 2011)	0.07116	0.4148	0.61888	0.36828	0.6761
		0.00932	0.00797	0.00623	0.36828	0.6761
	Change in paid-up capital (Mechanism 2018 vs 2011)	-0.0339	0.17192	0.18576	0.10791	0.58944
		0.00931	0.00659	0.0052	0.10791	0.58944
	Change in Lab. Productivity (Mechanism 2018 vs 2011)	-0.029	-0.3203	-0.2532	-0.2008	0.87841
		0.01516	0.00962	0.00547	-0.2008	0.87841
Board Characteristics	Reaching 2+ women (Learning 2018 vs. 2015)	0.1441	0.17516	0.28199	0.20041	0.40032
		0.00438	0.00474	0.00561	0.20041	0.40032
	Female-share of independent directors (2015)	0.55003	0.66166	0.8653	0.69233	0.97385
		0.01061	0.01209	0.01324	0.69233	0.97385
	0 or 1+ independent female directors (2015)	0.29217	0.34513	0.43762	0.35831	0.4628
		0.00545	0.00574	0.00596	0.35831	0.4628

-0.3
-0.2
-0.1
0
0.1
0.2
0.3

Position on standardised distribution

Figure 6: Means of covariates across subgroups

The color depicts where the mean value of a covariate in a subgroup lands on the standardised empirical distribution of its variable. Standard deviations of the subgroup-wise mean value in grey. These nine covariates are the top covariates, based on the highest difference in the standardised value between subgroup 1 and 3 (and similarly also found in the top 15 measured with equation (7)). Firm characteristics are average quarterly values for 2011 or 2018/2011. Labour productivity in log revenue per log wages, total wages in log, profits in m$, changes in terms of growth, i.e. (2018 value)/(2011 value).

Before jumping to any early conclusions, I check the robustness of the Causal Forest results by growing three new forests F1-F3, that test the most critical and sensitive assumptions - results in the figure 7. F0) Is the main forest as a reference, F1) a forest only with observations of 2020Q2/Q3 (4830 obs.) to test the insensitivity to the Panel time period, F2) is a forest with minimal observations per node set to 20 instead of 10 and F3) a forest built on the naive propensity (as introduced above).

The estimations of F1-F3 show that a) except for low-sample-size F1, all forests are correctly calibrated ($\upsilon = 1$), b) only the main forest can detect differences between subgroup one and three (p-values> 0.1), c) $\hat{\beta}^{forest}$ varies considerably between forests, though the $\beta = -0.595$

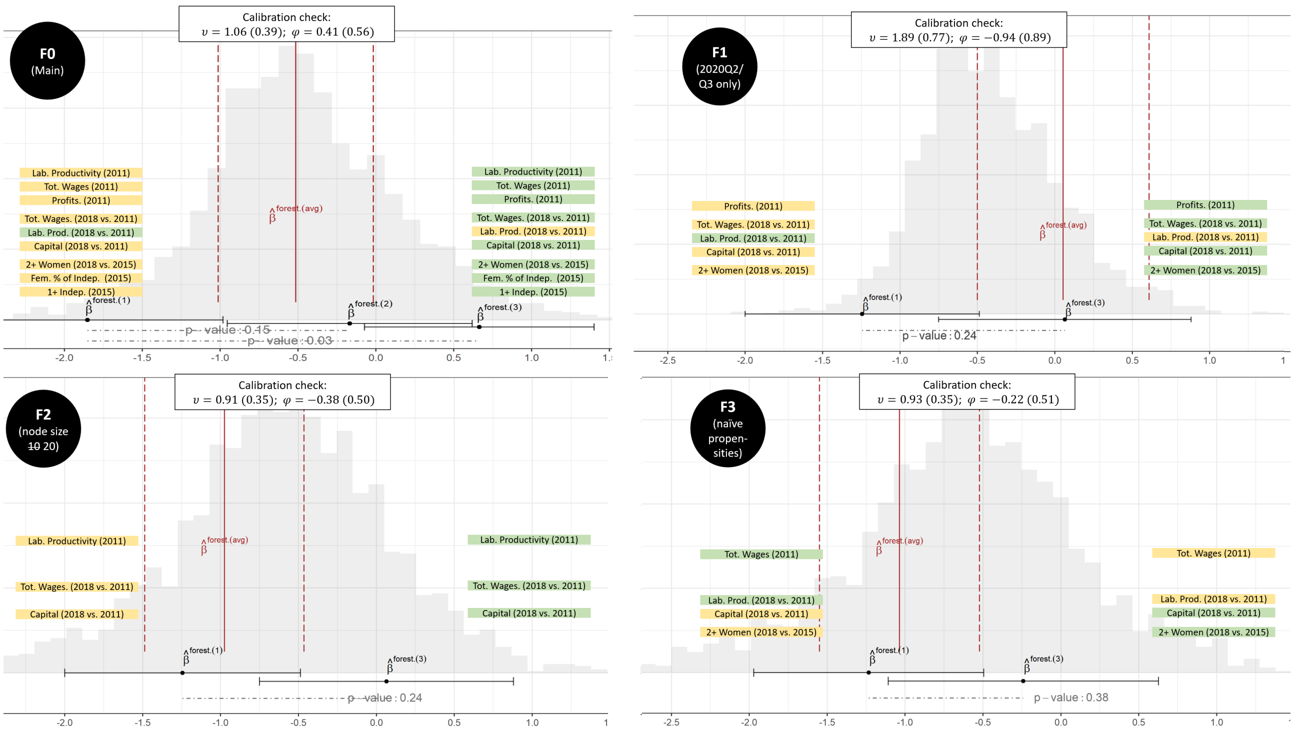

Figure 7: Robustness of the Causal Forest results

Calibration check parallel to Hyp. 4 above, negative results not meaningful. Reported p-value for subgroup 1 vs 3 (subgroup 2 is still there but not reported). Background colours for covariates in subgroup 1 (left) vs. 3 (right-hand side) simplified to yellow for lower end of standardised distribution and green for higher end, symmetric to figure above. For easier comparison, displayed covariates is the intersection of the 9 covariates from the main analysis F0 with the top 9 covariates from a given forest here

lies within one standard deviation of all $\hat{\beta}^{forest(avg)}$. There are good reasons why I use the setup F0 - and not F1 (low sample size), nor F2 (lower split precision), nor F3 (more confounded), but the variations in suggested covariates driving heterogeneity are still surprisingly big. At least, the relative subgroup-wise mean of a covariate (see colouring) is the same in almost all cases, including for covariates of less relevance for F1-F3[44], like the two independence covariates that are deemed relevant in F0. This finding is in line with the consistent finding in the literature on the positive effect of independent directors (e.g., Sarkar and Selarka (2021)). One covariate - capital increase before Covid-19 - is consistently high in the subgroup with the largest β^{forest}, which links well to crisis resilience theory that suggests higher capital makes firms less dependent on limited emergency loans, and is especially often used by risk-averse directors - most of them female (Palvia et al. (2015), Parrotta and Smith (2013),Dormady et al. (2019))

[44] Not shown her as they were not part of the top 9 for the forest, but are part of the top 20 (and subgroup-wise covariate means signif. different)

In the context of the broader female leadership literature, these mixed results are not surprising. As reviewed in the beginning, there has been evidence for positive, neutral and negative relationships between board gender diversity and financial performance. With the deeper heterogeneity analysis, I find firms in each of the three spectra. Instead of using average results, policy-makers and researchers might use methods like Causal Forest to understand heterogeneity for better targeted policies and structural improvements. For example, with policies fostering director independence[45]. However, Causal Forest used in highly confounded settings, like here for a heterogeneity-in-heterogeneities analysis, are limited to providing descriptive, initial indications of possible causalities. Two core challenges remain: 1) Identifying causal settings and 2) Gathering sufficient data for deconfounding and of interesting covariates for heterogeneities.

In summary, there are three main takeaways from the Causal Forest analysis. First, the detectable heterogeneity is limited, though there is an indication of a positive association between female leadership in the most-affected sectors and crisis resilience for one of the subgroups, and a negative for another subgroup. The latter subgroup exhibits relatively low growth in capital and wages (indicator of size) before Covid-19. Additionally, more independence for the board and adding more women might be beneficial for crisis resilience. Second, robustness checks show that one should be careful with interpretations of heterogeneity drivers when only one forest is estimated, and especially when the calibration check cannot reject the null of no heterogeneities. Third, even though the Causal Forest methodology was able to add additional granularity and point towards indicative channels, data limitations - low sample size and variable availability - are barriers to fully employ a deep heterogeneity analysis. This is ubiquitous in studies of very recent events. Causal Forest can potentially aid in disentangling a mass of covariate heterogeneities, but for this, there needs to be a mass of covariates. In complex settings with little data - like here - neither

[45] A positive example is the recent regulatory push for independent directors by the Securities and Exchange Board of India

traditional econometric nor modern Machine Learning methods are deeply insightful. This emphasises that complex gender settings can only be understood more deeply with the development of a) better data collecting or sharing systems and b) algorithms that work well with small samples.

4 Conclusion

This piece of research has revealed the value in a granular approach to answer corporate governance and female leadership questions. In summary, "When are large female-led firms more resilient against shocks?": Superficial Double Differences indicated that, on average, large Indian female-led firms experienced less strong revenue declines during Covid-19. Including the sector-wise magnitude of the shock, Triple Differences and Causal Forests suggest that this association dissipates in strongly affected sectors especially when firms are large and recent growth in labour productivity and capital has been relatively low.

The results presented here need be carefully interpreted and contextualised. First, the number of female directors is predicted with error. Second, the number of firms is relatively low, so that overlap is limited and larger neighbourhoods for comparisons along covariates are needed. Third, the definition of the sector-wise shock itself is influenced by the abilities and self-selection of women in specific sectors. Fourth, and most importantly, with a lack of data on pre-crisis and pre-policy resilience structures, as there is no recent crisis comparable to Covid-19 in which they might come to light, I cannot legitimately claim baseline balance. The allocation of female directors into firms is non-random, even though the policy might have led to a less endogenous selection than before. Fifth, the sample size cannot detect strong heterogeneities, leading to rather unrobust results on mechanisms behind covariates. When more data of more firms becomes available, future research might show stronger findings or could do improved matching of compliers and non-compliers.

Nevertheless, this piece of research has indicated the potential in combining Triple Differences and Causal Forests to analyse complex heterogeneities in the gender dimension of corporate leadership. Even though the full potential of Causal Forests is harnessed in carefully identified causal setups, the method can be valuable for observational data and analysis of heterogeneities within heterogeneities. Especially given the limited exogenous settings for gender analysis and other corporate governance dimensions, researchers either make strong assumptions to match observations of opposite gender, or often just add an interaction term of gender and a different treatment as a simple heterogeneity analysis. However, with enough data, there is potential in moving beyond this first stage heterogeneity analysis, towards a second stage of understanding heterogeneities within these heterogeneities. This can be relevant initial (non-causal) indications of structural (self-)selection and dynamics by gender - as shown with positive growth dynamics and independence in this analysis. Traditional econometric tools like regression interaction terms are not suitable for such a high dimensional analysis. Instead, researchers can harvest the advances in Machine Learning, connected to econometrics through Causal Forests. Future research could shed light on further twists to this methodology and use algorithms that work better in small-sample settings and gather data on more firms and covariates, beyond India and the gender dimension of corporate governance.

5 Appendix

5.1 Doubly Robust Propensity Score

The following paragraph provides a background for the use of the doubly robust propensity score in a causal forest context, orientated on Hirano and Imbens (2005), Austin (2019), Chernozhukov et al. (2018) and Athey et al. (2019). The continuously distributed treatment received, $T_i \in [t_0, t_1]$[46], and its corresponding potential outcome, $Y_i(t)$, are dependent on a vector of covariates X_i. By conditioning on this vector, weak unconfoundedness is assumed: $Y(t) \perp T|X$. Then, the generalised propensity score is $r(T, X)$ with the conditional density of treatment given covariates $r(t, x) = f_{T|X}(t|x)$. Then, for every t, $f_T(t|r(t, X), Y(t)) = f_T(t|r(t, X))$, i.e. the generalised propensity score is sufficient for weak unconfoundedness. Traditionally, the propensity score is used in a matching or an inverse propensity weighting procedure with $\frac{1}{f_{T|X}(t|x)}$ which is stable against infinite variance when adjusted to $\frac{f_T(t)}{f_{T|X}(t|x)}$[47]. Combining the propensity score model $e(x) = E(T|X = x) = f_{T|X}(t|x)$ with an outcome model $E(Y|X = x)$ is the basis of the doubly robust procedure. In the Causal Forest context, this is specifically achieved by predicting each with any non-parametric prediction technique, in the application here with random forests. With the learned $\hat{e}(x)$ and $\hat{m}(x)$, the inverse propensity is used for estimating the potential outcomes $Y_{it}^{Doubly} = \hat{m}(x) + \frac{Y_i - \hat{m}(x)}{\hat{e}(x)} * 1\{T_i = t\}$. Finally, for example for every level of sector-affectedness, β^{forest} is learned by regressing $Y_{it}^{Doubly} - Y_{i0}^{Doubly}$ on X_i with similarity weights. This doubly robust procedure has the two advantages that firstly, the effect estimate is only influenced by the product of $\hat{e}(x)$ and $\hat{m}(x)$ - as long as one of them is correct, the final model is correct. Secondly, under above described assumptions one to six, the final estimate is asymptotically normal and valid confidence intervals can be constructed.

[46] t here indicates a treatment level, in contrast to the use for time in the other subsections

[47] (With a normally distributed T with mean μ^T and variance σ^2, $f_{T|X}(t|x)$ can be estimated by the normal density $\frac{1}{2\pi\sigma^2} e^{-(\frac{t-\hat{\mu}x)^2}{2\hat{\sigma}^2}}$, $f_T(t) = \frac{1}{\sqrt{2\pi\hat{\sigma_t}^2}} e^{-\frac{(t-\mu_t)^2}{2\pi\hat{\sigma_t}^2}}$

5.2 Clustered augmented inverse propensity weighted treatment effects

With clustering of all observations it into groups per firm i (total firms $= n$), keeping in mind that the nuisance parameters are the same across clusters as X_i is constant per firm, the AIPW-based average treatment effects are computed as follows (based on Athey and Wager (2019) with slight adjustments). $\hat{\tau}$ is the equivalent of $\hat{\beta}^{forest}$:

$$\hat{\tau}_i = \frac{1}{T}\sum_{t=1}^{T}\hat{\Gamma}_{it}, \quad \hat{\tau} = \frac{1}{n}\sum_{i=1}^{n}\hat{\tau}_i, \quad \hat{\sigma}^2 = \frac{1}{n(n-1)}\sum_{i=1}^{n}(\hat{\tau}_i - \hat{\tau})^2,$$

$$\hat{\Gamma}_{it} = \hat{\tau}^{(-it)}(X_i) + \frac{\varepsilon_{it}^{WCT} - \hat{e}^{(-i)}(X_i)}{\hat{e}^{(-i)}(X_i)(1-\hat{e}^{(-i)}(X_i))}(\varepsilon_{it}^{Y} - \hat{m}^{(-i)}(X_i) - (\varepsilon_{it}^{WCT} - \hat{e}^{(-i)}(X_i))\hat{\tau}^{(-it)}(X_i))$$

References

Athey, S. and Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113(27):7353–7360.

Athey, S., Tibshirani, J., and Wager, S. (2019). Generalized random forests. *The Annals of Statistics*, 47(2).

Athey, S. and Wager, S. (2019). Estimating Treatment Effects with Causal Forests: An Application. *arXiv:1902.07409 [stat]*. arXiv: 1902.07409.

Athey, S. and Wager, S. (2021). Policy Learning With Observational Data. *Econometrica*, 89(1):133–161.

Austin, P. C. (2019). Assessing covariate balance when using the generalized propensity score with quantitative or continuous exposures. *Statistical Methods in Medical Research*, 28(5):1365–1377. Publisher: SAGE Publications Ltd STM.

Beaman, L. (2010). Political reservation and substantive representation: Evidence from Indian village councils.

Beaman, L., Duflo, E., Pande, R., and Topalova, P. (2012). Female Leadership Raises Aspirations and Educational Attainment for Girls: A Policy Experiment in India. *Science*, 335(6068):582–586.

Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1):5–32.

Brown, P. and Sung, J. (2015). *Higher Education, Corporate Talent and the Stratification of Knowledge Work in the Global Labour Market*. Routledge. Pages: 241-254 Publication Title: World Yearbook of Education 2015.

Burauel, P. and Schroeder, C. (2020). The German Minimum Wage and Wage Growth: Heterogeneous Treatment Effects Using Causal Forests. SSRN Scholarly Paper ID 3415479, Social Science Research Network, Rochester, NY.

Campbell, K. and Mínguez-Vera, A. (2008). Gender Diversity in the Boardroom and Firm Financial Performance. *Journal of Business Ethics*, 83(3):435–451.

Chattopadhyay, R. and Duflo, E. (2004). Women as Policy Makers: Evidence from a Randomized Policy Experiment in India. *Econometrica*, 72(5):1409–1443.

Chen, J., Leung, W. S., Song, W., and Goergen, M. (2019). Research: When Women Are on Boards, Male CEOs Are Less Overconfident. *Harvard Business Review*. Section: Boards.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68.

Coscieme, L., Fioramonti, L., Mortensen, L. F., Pickett, K. E., Kubiszewski, I., Lovins, H., Mcglade, J., Ragnarsdóttir, K. V., Roberts, D., Costanza, R., Vogli, R. D., and Wilkinson, R. (2020). Women in power: Female leadership and public health outcomes during the COVID-19 pandemic. *medRxiv*, page 2020.07.13.20152397. Publisher: Cold Spring Harbor Laboratory Press.

Dalal, H. (2020). Corporate India: WOmen on Boards. page 48.

Danes, S. M., Lee, J., Amarapurkar, S., Stafford, K., Haynes, G., and Brewton, K. E. (2009). Determinants of family business resilience after a natural disaster by gender of business owner. *Journal of Developmental Entrepreneurship*, 14(04):333–354. Publisher: World Scientific Publishing Co.

Deloitte (2018). Women in the Boardroom | Deloitte | Press Release.

Dormady, N., Roa-Henriquez, A., and Rose, A. (2019). Economic resilience of the firm: A production theory approach. *International Journal of Production Economics*, 208:446–460.

Efron, B. (2014). Estimation and Accuracy After Model Selection. *Journal of the American Statistical Association*, 109(507):991–1007.

Farrell, K. A. and Hersch, P. L. (2005). Additions to corporate boards: the effect of gender. *Journal of Corporate Finance*, 11(1):85–106.

Ford, D., P, R., and Ford, D. (2012). *Gender Quotas and Female Leadership: A Review. Background Paper for the World Development Report on Gender*.

Francoeur, C., Labelle, R., and Sinclair-Desgagné, B. (2008). Gender Diversity in Corporate Governance and Top Management. *Journal of Business Ethics*, 81(1):83–95.

Funk, M. J., Westreich, D., Wiesen, C., Stürmer, T., Brookhart, M. A., and Davidian, M. (2011). Doubly Robust Estimation of Causal Effects. *American Journal of Epidemiology*, 173(7):761–767.

Gutiérrez-Fernández, M. and Fernández-Torres, Y. (2020). Does Gender Diversity Influence Business Efficiency? An Analysis from the Social Perspective of CSR. *Sustainability*, 12(9):3865. Number: 9 Publisher: Multidisciplinary Digital Publishing Institute.

Hillman, A. J., Withers, M. C., and Collins, B. J. (2009). Resource Dependence Theory: A Review. *Journal of Management*, 35(6):1404–1427. Publisher: SAGE Publications Inc.

Hirano, K. and Imbens, G. W. (2005). The Propensity Score with Continuous Treatments. In Gelman, A. and Meng, X.-L., editors, *Wiley Series in Probability and Statistics*, pages 73–84. John Wiley & Sons, Ltd, Chichester, UK.

Imbens, G. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika*, 87(3):706–710.

Kennedy, E. H., Ma, Z., McHugh, M. D., and Small, D. S. (2017). Non-parametric methods for doubly robust estimation of continuous treatment effects. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(4):1229–1245.

Lechner, M. (2019). Modified Causal Forests for Estimating Heterogeneous Causal Effects. *arXiv:1812.09487 [econ, stat]*. arXiv: 1812.09487.

Lee, A., Inceoglu, I., Hauser, O., and Greene, M. (2020). Determining causal relationships in leadership research using Machine Learning: The powerful synergy of experiments and data science. *The Leadership Quarterly*, page 101426.

Liu, Y., Wei, S., and Xu, J. (2021). COVID-19 and Women-Led Businesses around the World. *Finance Research Letters*, page 102012.

Mikkola, A. (2005). Role of Gender Equality in Development - a Literature Review. SSRN Scholarly Paper ID 871461, Social Science Research Network, Rochester, NY.

Miller, S. (2020). Causal forest estimation of heterogeneous and time-varying environmental policy effects. *Journal of Environmental Economics and Management*, 103:102337.

Morrison, A., Raju, D., and Sinha, N. (2007). *Gender Equality, Poverty and Economic Growth*. World Bank Publications. Google-Books-ID: 2ch8Sh63xggC.

Nguyen, T., Locke, S., and Reddy, K. (2015). Does Boardroom Gender Diversity Matter? Evidence from a Transitional Economy.

O'Neill, E. and Weeks, M. (2018). Causal Tree Estimation of Heterogeneous Household Response to Time-Of-Use Electricity Pricing Schemes. Working Paper, Faculty of Economics. Accepted: 2018-12-07T15:00:37Z.

Palvia, A., Vähämaa, E., and Vähämaa, S. (2015). Are Female CEOs and Chairwomen More

Conservative and Risk Averse? Evidence from the Banking Industry During the Financial Crisis. *Journal of Business Ethics*, 131(3):577–594. Publisher: Springer.

Park, A., Yang, D., Shi, X., and Jiang, Y. (2009). Exporting and Firm Performance: Chinese Exporters and the Asian Financial Crisis. Technical Report w14632, National Bureau of Economic Research.

Parrotta, P. and Smith, N. (2013). Female-Led Firms: Performance and Risk Attitudes. SSRN Scholarly Paper ID 2329083, Social Science Research Network, Rochester, NY.

Rebérioux, A. and Roudaut, G. (2019). The Role of Rookie Female Directors in a Post-Quota Period: Gender Inequalities within French Boards. *Industrial Relations: A Journal of Economy and Society*, 58(3):423–483. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/irel.12238.

Reddy, S. and Jadhav, A. M. (2019). Gender diversity in boardrooms – A literature review. *Cogent Economics & Finance*, 7(1):1644703. Publisher: Cogent OA _eprint: https://doi.org/10.1080/23322039.2019.1644703.

Rose, C. (2007). Does female board representation influence firm performance? The Danish evidence. *Corporate Governance: An International Review*, 15(2):404–413. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-8683.2007.00570.x.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701. Place: US Publisher: American Psychological Association.

Sarkar, J. and Selarka, E. (2021). Women on board and performance of family firms: Evidence from India. *Emerging Markets Review*, 46:100770.

Sen, A. K. (1999). *Development as freedom*. Knopf, New York, 1st. ed. edition. Open Library ID: OL13625892M.

Singh, A., Singhania, S., and Sardana, V. (2019). Do Women on Boards affect Firm's Financial Performance? Evidence from Indian IPO Firms. *Australasian Accounting, Business and Finance Journal*, 13(2):53–68.

Srivastava, V., Das, N., and Pattanayak, J. K. (2018). Women on boards in India: a need or tokenism? *Management Decision*, 56(8):1769–1786. Publisher: Emerald Publishing Limited.

Terjesen, S., Couto, E. B., and Francisco, P. M. (2016). Does the presence of independent and female directors impact firm performance? A multi-country study of board diversity. *Journal of Management & Governance*, 20(3):447–483.

Torres, A. P., Marshall, M. I., and Sydnor, S. (2019). Does social capital pay off? The case of small business resilience after Hurricane Katrina. *Journal of Contingencies and Crisis Management*, 27(2):168–181. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1468-5973.12248.

Vieira, E. S. (2017). Board of directors characteristics and performance in family firms and under the crisis. *Corporate Governance: The International Journal of Business in Society*, 18(1):119–142. Publisher: Emerald Publishing Limited.

Wager, Athey, and Tibshirani, J. (2020). Generalized Random Forests [R package grf version 1.2.0]. Publisher: Comprehensive R Archive Network (CRAN).

Wager, S. and Athey, S. (2018). Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. *Journal of the American Statistical Association*, 113(523):1228–1242. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/01621459.2017.1319839.

Wang, M. and Kelan, E. (2013). The Gender Quota and Female Leadership: Effects of the Norwegian Gender Quota on Board Chairs and CEOs. *Journal of Business Ethics*, 117(3):449–466.